Xjenza Online - Journal of The Malta Chamber of Scientists
www.xjenza.org
DOI: 10.7423/XJENZA.2014.2.09

Research Article

Representational Momentum and the Human Face: an empirical note

Ian M. Thornton
Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Msida, MSD 2080, Malta.

Abstract. Recent evidence suggests that observers may anticipate the future emotional state of an actor when viewing dynamic expressions of emotion, consistent with the notion of representational momentum. The current paper presents data that conflicts with these previous studies, finding instead that memory for the final frame of an emotional video tends to be shifted back in the direction of the first frame. While simple methodological issues may explain this difference (e.g., the use of morph sequences in previous studies versus naturalistic expressions here) a more theoretically interesting possibility is also considered. Specifically, recent studies of ensemble representations have shown that observes can rapidly extract the average expression from a display of up to 20 faces. It is suggested that the need to predict versus the need to maintain a stable estimate of the current state often compete when we interact with dynamic stimuli. Our memory for the final expression on an emotional face may be particularly sensitive to task demands and response timing, thus coming to reflect different solutions to this anticipation-averaging conflict depending on the precise experimental scenario.

Keywords Representational Momentum – Dynamic Faces – Ensemble Representations – Anticipation – Averaging – Emotional Expressions

1 Introduction

Representational momentum (RM) refers to a perceptual phenomenon in which observers tend to misremember the stopping point of a dynamic event as being further forward in its direction of motion or change (Freyd and Finke, 1984; Freyd, 1987). In a now classic example, reproduced in Figure 1, three views of a rectangle were presented sequentially so as to imply rotation in the picture plane (Freyd and Finke, 1984). After this inducing sequence, the screen went blank for a brief retention interval and a probe rectangle appeared. Across trials, the probe rectangle was presented either in the true/same orientation as the last inducing image or in positions parametrically varied forward or backward from this orientation. Participants simply indicated whether the presented probe was the same or different relative to the remembered stopping point. As can be seen in Figure 1B, observers were much more likely to endorse probes as being the "same" as the stopping point when those probes were shifted forward in the direction of implied rotation.

Similar "forward shifts" have been reported for a wide range of transformations and a number of mechanisms have been proposed to explain the phenomenon (for review, see Hubbard, 2005). Important in the current context are findings that show that RM is not restricted to simple displays in which a single geometric figure undergoes translation or rotation. Specifically, forward shifts have been found in displays containing multiple elements (Finke and Shyi, 1988; Thornton and Hayes, 2004) and for complex transformations, such as those involving the articulating human form (Graf et al., 2007; Verfaillie and Daems, 2002; Verfaillie et al., 1994). Although there is continued debate about the nature of RM shifts, it is possible that their general role is to help anticipate or predict future states. The goal of the current paper is to comment on recent findings suggesting similar anticipatory processes may occur when we view dynamic faces, possibly helping us to predict the future emotional state of others (Marian and Shimamura, 2013; Palumbo and Jellema, 2013; Uono et al., 2009, 2014; Yoshikawa and Sato, 2008).

Faces are fascinating objects. Although they all share

Correspondence to: Ian M. Thornton (ian.thornton@um.edu.mt)

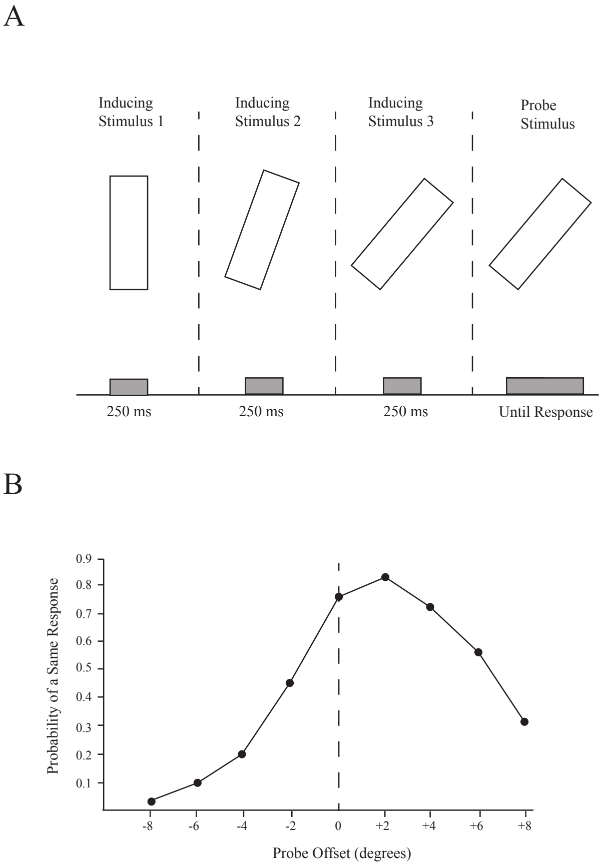

Figure 1: Display and data from Freyd and Finke (1984), redrawn from original. Panel A shows the inducing display and probe items. The three successive views of the rectangle were separated by a blank inter-stimulus interval (ISI) of 250 ms and a retention interval (RI) of 250 ms (all shown as dashed vertical lines). The probe item was either the same as the final inducing item or was varied between −8 and 8 degrees around the true final orientation. The task was to indicate via a key press whether the probe had the same or different orientation. Panel B shows data from 16 participants. It is clear that participants responded "same" more frequently to probes shifted in the direction of motion. Such a shift gave rise to the term "representational momentum".

the same basic configuration – eyes above nose, above mouth – subtle variations in the arrangement of these features make it possible to produce the huge variety of patterns that form the basis of our individual identity. Beyond this variability in configuration, differences in skin texture, 3D shape and, most relevant here, motion, further enhance identity and also make it possible for the face to convey information about race, gender, age, speech and emotion. For a recent, comprehensive review of face research, see Bruce and Young (2012). While the majority of work in face processing literature has employed static, photographic stimuli, the last 20 years has also seen a rapid growth of research into the role of facial motion. Work from my own group, for example, has consistently shown that when we become familiar with a new individual, we use their characteristic facial motion, as well as their facial shape, to encode identity (Knappmeyer et al., 2003; Pilz et al., 2005; Thornton and Kourtzi, 2002). For a general overview of research in this area see (O'Toole et al., 2002; O'Toole and Roark, 2010; Xiao et al., 2014).

Within the context of moving faces, Yoshikawa and Sato asked whether observers would misremember the stopping point of a dynamic face showing emotion (Yoshikawa and Sato, 2008). Using a standard set of photographic stimuli (Ekman and Friesen, 1976), they constructed linear morphs between a neutral face and 6 basic expressions of emotion. On each trial a morph was shown as a brief movie, and stopped at either 80, 90 or 100 % of the peak of the expression. After a brief, blank pause, a probe item appeared and the observer was allowed to adjust the movie frame to find the remembered stopping point. Particularly when movies were played faster than normal and when the true stopping point was around 80 % of the peak of the expression, (Yoshikawa and Sato, 2008) found clear evidence for RM. That is, observers consistently selected a frame that was beyond the actual stopping point of the animation.

The aim of this brief commentary is not to directly question this basic finding. Indeed this RM effect has been replicated several times (Uono et al., 2009, 2014) and patterns consistent with anticipation have also been reported in other labs (Marian and Shimamura, 2013; Palumbo and Jellema, 2013). Rather, the goal here is to raise a note of caution in generalising from these findings and to suggest that when processing dynamic faces, more than one mechanism may be influencing behaviour. The "empirical" part of the title to this paper relates to previously unpublished data from my doctoral dissertation, supervised by Jennifer Freyd at the University of Oregon, that bear directly on the question of RM and faces (Thornton, 1997). The next section provides a brief overview of the methods used to study RM in this thesis and presents some of the findings. In short, over a number of experiments, "backward" rather than "forward" shifts were always obtained. That is, observers typically endorsed probes that were earlier not later in the sequence they had seen. The final section of the paper will relate these findings to previous studies that have shown anticipatory effects, and discuss what factors may account for such different results.

2 Experimental Methods and Results

This section will briefly describe some of the experiments reported in Chapters V-VI of Thornton (1997) as well as some additional control experiments. Figure

2 shows a 20-frame expression sequence typical of those used to explore RM in Thornton (1997). Although having access to the same expressive photographs (Ekman and Friesen, 1976) used in Yoshikawa and Sato (2008), it was decided instead to use naturalistic expressions. An elicitation technique was developed in which volunteers performed a number of tasks, such as recognising sounds, exploring objects by touch, watching comedy footage, all while being filmed. From these filmed sessions, frame sequences of facial expressions, such as smiles, frowns and looks of disgust were extracted to provide both the RM inducing sequences and the probe items. For more details on the elicitation technique and stimuli see Chapter II of Thornton (1997) or Thornton and Kourtzi (2002).

Figure 2: The 20 frame sequence used in Experiments 5 and 7 of Thornton (1997). Numerals refer to frame numbers. To create an inducing sequence, the frames were played sequentially at a rate of 30 fps as an aperture movie with a frame that subtended approximately 3° visual angle. The movie always stopped on frame 10, which is drawn here with a exaggerated border, not shown during the experiment. In Experiment 5, the critical inducing sequence was termed "towards expression" (TE) and involved frames 1-10, so that the expression was played in the normal direction. In Experiment 7, the inducing sequence was termed "towards neutral" (TN) and consisted of frames 20-10, the expression thus unfolding in a reverse direction. In these original experiments, the inducing sequences were also preceded by a longer preview animation in which all frames were shown once in both directions. Thus, for the TE condition the preview consisted of frames 1-20 followed by frames 20-1. The entire movie thus consisted of 50 frames and had a duration of 1500 ms. See text for details on the rationale and possible consequences of this preview design.

During each trial, a sequence of frames was played at normal speed (30 fps) and then stopped abruptly at a given point. Across experimental conditions, the movie could be playing forward (frames 1-10, neutral-to-smile in Figure 3) or in reverse (frames 20-10, smile-to-neutral). Importantly, across conditions the stopping point was always the same (frame 10 in this example). The screen remained blank for a 300 ms retention interval and then a probe item appeared. Figure 3 shows examples of the probe items used for the sequence in Figure 2. The probe item could be identical with the true/same stopping point (Probe 0), or was parametrically varied around that point, one or two steps forward of the stopping point or one or two steps back. Note how, across conditions, the same frames appear as different probe items, depending on the direction of the animation, a point I return to shortly. In all trials, the observer simply had to respond whether the probe was the same or different from the point at which the movie stopped. This is essentially the same experimental design used in Freyd and Finke (1984) and illustrated in Figure 1.

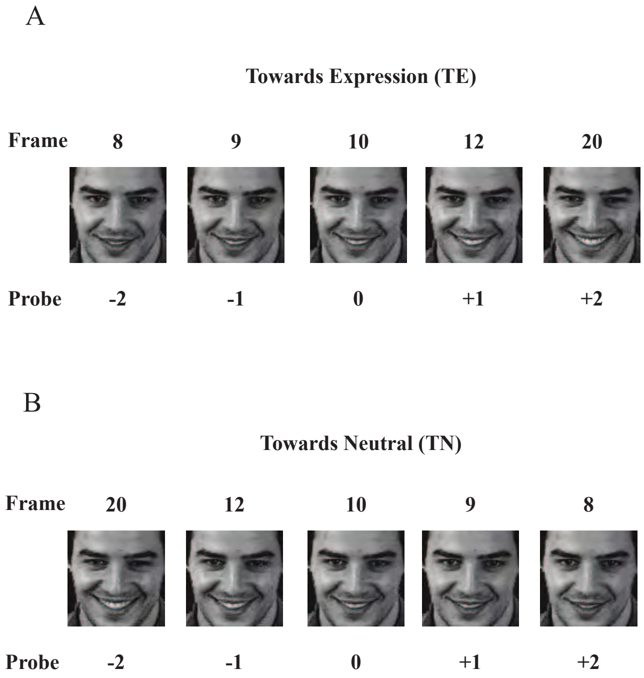

Figure 3: The probe images used in Experiments 5 (Panel A) and 7 (Panel B) of Thornton (1997). The probe always appeared after a 250 ms blank retention interval that followed the inducing sequence. The probe remained visible until participants made a "same" or "different" response using the standard keyboard. Note that the same images appear in both TE and TN conditions, but are assigned different probe positions due to the direction of the inducing sequence. The probe spacing was determined using separate concurrent matching tasks, and attempts to reflect equal perceptual spacing rather than frame spacing per se. For details of the concurrent matching task see Thornton (1997).

As noted in Yoshikawa and Sato (2008), there are a number of issues that need to be addressed when using naturalistic footage, rather than artificial, controlled stimuli to explore RM. The most serious concern is the selection of appropriate probe items. If there is an asymmetry in the amount of change in frames before or after the true stopping point, then this probe asymmetry, rather than the display dynamics could cause a shift. We tried to control for this in two ways. First, we ran separate concurrent matching experiments to determine the distance in frames that needed to be used in matched pairs of probes in order to achieve equal discriminability (i.e., probes $1/-1$ and $2/-2$ in Figure 3). See Thornton (1997) for more details on how this matching task was used objectively to measure the amount of perceived change in each probe item. More importantly, by running each inducing sequence in both forward and reverse directions we could disentangle shifts that arose due to the dynamic inducing sequence from those that depended on probe discriminability. For example, with reference to Figure 3, if the slightly more intense smile (frame 12) was harder to discriminate from the stopping point than the less intense smile (frame 8) then this would lead to a forward shift in one condition and a backward shift in the other.

Example results are shown in Figure 4. Panel A shows data from 26 participants who were shown the male sequence in Figure 2. Twelve participants completed the "towards expression" (TE) condition and 14 the "towards neutral" (TN) condition. Each participant completed 100 experimental trials, 20 repetitions at each probe position, presented in a random order. These data were reported as Experiments 5 and 7 in Thornton (1997). Panel B shows data from a within-subjects experiment (N =12) that has the same design, except the same participants completed both TE and TN conditions in separate, counterbalanced blocks. This experiment, which used a female face frowning, was run as a control condition but was not included in the thesis.

There are two important points to note. First, there is no evidence of a forward shift in any of the four conditions. Rather, participants were always more likely to endorse as same a probe that matched an earlier (probe -1) rather than a later (probe 1) point in the sequence. This led to a consistent backward rather than a forward shift. To quantify the overall pattern of responses a weighted mean (WM) was calculated on the distribution of each participant's data. The WM is calculated by multiplying the proportion of 'same' responses at each probe position by that probe's distance from the true/same probe (i.e., 0). The sum of products is then divided by the total number of 'same' responses to yield an appropriately weighted measure of central tendency (Faust, 1990; Gray and Thornton, 2001). All 4 of the conditions in Figure 4 A-B has a weighted mean that is reliably less than zero according to one-sample t-tests (all $ps < 0.05$). See Table 1 for further details.

The second point to note is that individual probe items do not appear to be influencing the pattern of results. As mentioned above, across the two conditions the exact same image changes its position relative to the true stopping point as a function of the direction of the animation. So, probe 1 in the TE condition (frame 12) becomes probe -1 in the TN condition, for example. If responses depended on asymmetric probe discriminability, the two curves in each panel would be mirror images of each other, rather than following the same basic shape. The fact that the two curves do not completely overlap illustrates the way in which the dynamic context of the inducing sequences affects the interpretation of the same probe images (see also Palumbo and Jellema, 2013). For example, in Figure 4A, it seems likely that the rate of expressive change in the TE sequence (frames 1-10) is greater than in the TN sequence (frames 20-10) leading to different patterns of shifts. The slight difference in duration between the two conditions – the TE sequence actually consists of 10 frames, the TN has 11 frames – could also be a factor, although this seems unlikely as the opposite TE-TN pattern occurs in Figure 4B. These asymmetries, then, seem more likely to reflect idiosyncratic expressive changes, a consequence of using naturalistic expression sequences. Again, however, it is important to stress that none of these shifts show any hint of anticipation.

To further establish that the presence of negative shifts was not an artefact of probe selection, we ran a control condition with linear, schematic faces. Stimuli consisted of an oval face in which two small circles for eyes and a plus sign for the nose were centred and colinear. Three identical lines formed the two eyebrows and the mouth. These lines were expanded and contracted in a linear fashion to convey an impression of surprise. As in previous morph studies (Uono et al., 2009, 2014; Yoshikawa and Sato, 2008) probe items were guaranteed to be equidistant from the true stopping point. The data from the TE condition (reported as Experiment 5 in Thornton, 1997) are shown in Figure 4C. Again, the average weighted mean for this dataset was reliably less than zero.

It should be noted that in all of the experiments reported so far in this paper, the inducing sequence on each trial was preceded by a longer preview animation. Specifically, for the TE condition shown in Figure 2, all frames were played forward (frames 1-20), then backward (frames 20-1) before the final inducing sequence of frames 1-10. This was done to ensure that observers experienced the full range of the expression and had been exposed to probe items both before and after the stopping point. Although such preview animations are not typical of RM studies involving simple rotation or translation, it was felt that the complexity of the nonrigid facial transformations warranted providing additional contextual support. Nevertheless, such context could conceivably have influenced the pattern of shifts in

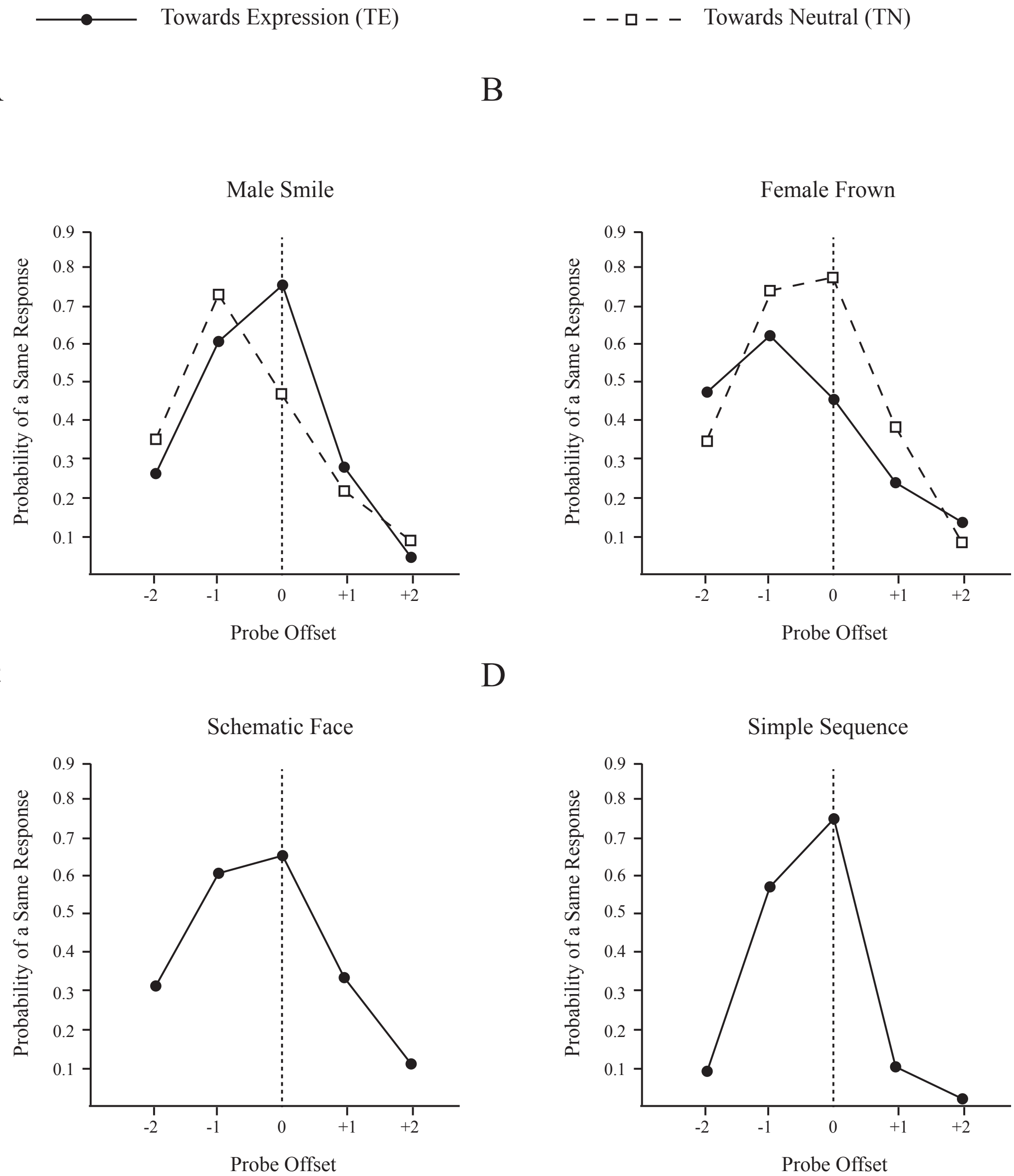

Figure 4: Experimental Results. A) Naturalistic Male Smile. Results from between subject design using data from Experiments 5 (Solid Line) and 7 (Dashed Line) from Thornton (1997). In contrast to the pattern shown in Figure 1B, the peak of the distribution of same responses in both these experiments in shifted backward, giving rise to negative weighted means (see Table 1). B) Naturalistic Female Frown. Results of a within-subjects control experiment using the same experimental design but different facial sequences. C) Schematic Face. Results using a simple, schematic face from Experiment 6, Thornton (1997). D) Simplified Inducing Sequence. Results of a control experiment where the inducing sequence in Figure 2 was simplified to consist only of frames 1-10.

some way. To explore this possibility, a further control experiment was conducted (some years after the thesis) using the same experimental stimuli but with the preview context removed. Figure 4D shows data from this experiment that used an identical design except that a brief, unidirectional expression sequence (frames 1-10 in Figure 2) was used as the inducing display. It is clear, that even without the more complex sequence, there is still a strong negative shift with a weighted mean reliably less than zero.

3 Discussion

The results described above tell a very simple story. Using naturalistic stimuli and a method very similar to that used in classic representational momentum work, there was no evidence for any form of forward, anticipa-

Table 1: Weighted Means (WM) for datasets in Figure 4.

Data Set	Condition	WM	t-value	p-value
Experiment 5 (Figure 4A)	TE	−0.36	$t(11) = 3.8$	$p < 0.01$
Experiment 7 (Figure 4A)	TN	−0.54	$t(13) = 6.7$	$p < 0.001$
Figure 4B	TE	−0.57	$t(11) = 2.9$	$p < 0.05$
Figure 4B	TN	−0.38	$t(11) = 2.8$	$p < 0.05$
Experiment 6 (Figure 4C)	TE	−0.34	$t(11) = 4.7$	$p < 0.001$
Figure 4D	TE	−0.38	$t(11) = 7.6$	$p < 0.001$

tory shift. Rather, and in direct contrast to more recent literature, it seems that observers in these experiments misremembered the facial motion sequence as stopping earlier than was actually shown. How can we explain this different pattern of results and, more importantly, what might it mean for our understanding of face perception in the context of motion?

It remains a possibility that the use of naturalistic stimuli introduced some issues with the probe items that were selected, and this influenced the pattern of results. However, finding the same shifts regardless of whether the inducing sequence was played forward or backward – meaning that the role of individual probe items were also reversed – would seem to argue against this explanation. The use of naturalistic stimuli also means that the movements themselves were more complex and almost certainly less predictable than the linear morphs used in studies that have shown anticipatory effects (Marian and Shimamura, 2013; Palumbo and Jellema, 2013; Uono et al., 2009, 2014; Yoshikawa and Sato, 2008). Again, though, the use of an extended context meant that observers were shown the range of movements possible on each trial, and would quickly have become highly familiar with the facial actions involved. Perhaps more compellingly, the same pattern of negative shifts was observed in Thornton (1997) even when the stimuli consisted of a schematic face where the movements were simple, linear and highly predictable. A more recent study by Courgeon et al. (2010) used an animated virtual character to explore RM shifts. These animations involved more complex movement than a linear morph, but still provided precise control over probe separation. The pattern of results in this study were largely consistent with Thornton (1997), that is, shifts were mainly backward rather than forward.

In terms of methodology, it seems more likely that differences in the task used to assess the shift might be the crucial factor. As reviewed in Hubbard (2005), across the many studies examining RM, there have been two typical methods, one involving probe items, psychophysically equivalent to the method of constant stimuli, and one involving direct response methods, such as positioning a cursor, psychophysically equivalent to the method of adjustment. In the current context, studies that have reported backward shifts (Courgeon et al., 2010; Thornton, 1997) appear to have used the former category of assessment while those reporting forward shifts (Uono et al., 2009, 2014; Yoshikawa and Sato, 2008), the latter. Importantly, at least for translating stimuli, these two methods have previously been shown to yield different results, with assessment by probe items giving rise to smaller shifts than direct pointing methods (Kerzel, 2003a). It would clearly be very interesting to take the same set of facial stimuli and measure performance using these two different methods.

Recently, Jellema and colleagues have argued that top-down "emotional anticipation", rather than low-level visual mechanisms, may explain dynamic distortions when interpreting facial expressions (Jellema et al., 2011; Palumbo and Jellema, 2013). In their work, participants rated the perceived emotion of a neutral image that was preceded by an animation moving from a peak expression towards neutral. For example, following a sequence moving from angry towards neutral, the neutral probe was rated slightly happier, an effect termed a perceptual "over-shoot" bias (Palumbo and Jellema, 2013).

The emotional anticipation explanation suggests that when we view a dynamic face, we may "involuntarily keep track of the emotional state of mind of the actor and implicitly anticipate what the actor's emotional state would be in the immediate future" (Palumbo and Jellema, 2013, p. 10). Interestingly, both "over-shoot" and "under-shoot" effects were observed by Palumbo and Jellema (2013) depending on whether the preceding animation was unidirectional or looped. This further indicates that the precise dynamic context can play an important role in face perception, and may be one factor that explains the negative shifts of Thornton (1997).

Another recent paper, Marian and Shimamura (2013), used a method very similar to Palumbo and Jellema (2013) to explore both contrast effects (i.e. a neutral face appearing happy after an angry-neutral animation) and momentum effects (i.e. an animation stopping at a 50 % happy face appearing happier). Across several experiments, these authors always found patterns consistent with a forward or "over-shoot" effect, rather than an "under-shoot" or negative shift as found in Thornton (1997). While they note that a single underlying mechanism could account for both contrast and momentum effects, they also point out that the latter seems particularly sensitive to facial context. Specifically, different patterns where found for angry-happy than for sad-happy dimensions and no momentum effect was observed when the stopping point was beyond 75 % maximum. This sensitivity could again help to explain the lack of momentum effects in Thornton (1997), given the use of relatively complex, naturalistic sequences in those experiments.

More generally, it is important to stress again that all of the studies reporting facial anticipation have used the same linear morphing technique to create their stimuli. Indeed, with the exception of Marian and Shimamura (2013), who used the NimStim set of expressions (Tottenham et al., 2009), all of the experiments based their morphs on the same photographs (Ekman and Friesen, 1976). It would certainly be very interesting to establish whether anticipation effects can be observed for a broader class of facial stimuli, particularly containing naturalistic movements rather than linear morphs.

As already mentioned, while morph stimuli may appear subjectively "natural" (Sato and Yoshikawa, 2004), they clearly involve less complex movement than we typically encounter with a face. For example, naturalistic human smiles have specific onset, apex and offset timings, which result in non-linear acceleration/deceleration profiles (Krumhuber and Kappas, 2005; Schmidt et al., 2003). Observers are sensitive to variations in the timing of these components (Edwards, 1998) and may use them to infer the emotional state of an actor (Krumhuber and Kappas, 2005). It is unclear if the absence of such profiles affects the RM task, but it is conceivable that linear inducing sequences and probe items significantly reduce the complexity of anticipation. While the issue of probe selection with naturalistic sequences was certainly not dealt with perfectly in Thornton (1997), the use of either perceptual normalisation or image processing techniques to equate task-relevant differences between probes, should make it possible to move away from strictly linear morphs with some of these designs.

So far, a number of methodological issues have been mentioned to try and account for the appearance of backward shifts in Thornton (1997). Perhaps a more important, or at least more theoretically interesting, question is to ask "why" a backward shift might ever occur? The idea of anticipation, emotional or perceptual, provides a clear rationale for expecting a forward shift: helping an observer predict and respond to what will happen next. Could there also be mechanisms at work that would lead to a backward shift? Would they have any functional significance?

In the broader RM literature, there have been reports of backward shifts in a number of previous papers. For example, backward shifts have been found when target stimuli change their perceived depth (Hubbard, 1996), auditory pitch (Hubbard, 1993a; Johnston and Jones, 2006) or luminance (Brehaut and Tipper, 1996). Similar backward shifts have also been reported in displays with more than one object (Hubbard, 2005, 1993b; Kerzel, 2003b) or tasks involving "bounded" changes, either due to periodic motion or physical barriers of some description (Hubbard and Bharucha, 1988; Verfaillie and D'Ydewalle, 1991). As discussed by Hubbard (2005), these patterns have generally been taken as an indication that factors other than RM are influencing behaviour in a given task. While a number of specific factors have been proposed, most relevant to the current discussion are the ideas of memory averaging and boundary conditions (Hubbard, 2005).

In his review paper, Hubbard (2005) suggested that the patterns described in Thornton (1997) occurred either because the nature of the facial changes were too complex to anticipate or were bounded by the physical limits of facial deformation. While the former point has already been discussed, the latter misses the nature of the sequences used in Thornton (1997). Hubbard states that "there is a clear maximum extent of facial expression, and the observers presumably anticipated a return to a more neutral expression, rather than an increasing intensity of the current expression beyond the maximum biological extent" (Hubbard, 2005, p. 831). However, as can be seen in Figure 2, the endpoint of the inducing sequences were specifically chosen to be away from expression endpoints, and were also presented both towards expression and towards neutral, reducing the likelihood of a simple boundary effect. There is, nevertheless, a more subtle way that performance could have been "bounded". That is, although the facial transformations themselves were continuous, perceptually, a number of factors might differentially weight parts of the sequence. For example, the endpoints might exert some form of anchoring effect (Marian and Shimamura, 2013; Palumbo and Jellema, 2013; Russell and Fehr, 1987), there may be perceptual discontinuities around expression boundaries (de Gelder et al., 1997; Etcoff and Magee, 1992) or the resolution of expression processing may be limited, leading to "perceptual key frames" of some sort. Any of these factors could give rise to "attractor" expressions that might disrupt simple RM effects. Clearly, these are several possibilities that could be explored in future studies.

Thornton (1997) speculated that some form of memory averaging (Freyd and Johnson, 1987) might be the most likely cause of the observed pattern of backward shifts. An early study by Vicki Bruce and colleagues (Bruce et al., 1991) has shown that over the course of an experiment, participants appeared to extract the average configuration of a set of facial images in which the features had been slightly adjusted, similar to the prototype effects of Posner and Keele (1968). However, the time course of this prototype effect seemed at odds with the supposed trial-by-trial effects of RM.

Over the last decade, however, there has been a growing interest in the idea of rapid ensemble representations (see Alvarez, 2011 for review). That is, there is now a great deal of evidence that the visual system can very quickly (e.g., < 1 second) extract summary representations along a number of dimensions, when presented with complex displays, representations that provide the observer with a precise estimates of central tendency. For example, when briefly shown a spatial array containing 16 dots of various diameters, observers immediately have access to information about the average size (Ariely, 2001). Similar effects have been shown for other basic feature dimensions, such as the mean orientation, motion and brightness of large sets of items (Alvarez, 2011). The general suggestion is that such computations could provide a very efficient means to reduce perceptual bandwidth by providing fairly precise estimates of a set without the need to visually inspect every member.

Importantly in the current context, these averaging mechanisms also seem to be at work when we process faces. Jason Haberman, David Whitney and colleagues, for example, have shown that when observers are presented with an array containing multiple faces, they are able to rapidly extract average emotion and gender (Haberman et al., 2009; Haberman and Whitney, 2007, 2009). Work from other labs has show similar findings for identity (de Fockert and Wolfenstein, 2009) and race (Srismith et al., 2014). Of most relevance here, this averaging process for faces also appears to work over time (Haberman et al., 2009). In this study observers were shown animations containing up to 20 frames from morphed expression sequences. As with the spatial arrays used in previous work (Haberman and Whitney, 2007; Haberman et al., 2009) observers were able to accurately estimate the average expression in these displays. It is important to note that the frames making up the animations in Haberman and Whitney (2009) were presented in a totally random order, not as a coherent expression change. Nevertheless, if ensemble representations do have any functional role to play in the everyday perception of faces, we would clearly expect averages to be computed when normal expression sequences are shown.

Could such estimates of average expression help explain the "backward" shifts observed in Thornton (1997)? Given the directional nature of the inducing sequences used in those experiments, it does appear that the "average" should fall somewhere back along the sequence between the first and final frame. Thus, in contrast to an RM prediction, this might favour some form of backward shift. It seems unlikely that observers in these studies would be directly confusing the true stopping point with the sequence average. Nevertheless, if such an average is automatically computed during the presentation of the inducing sequence, its existence might attract or otherwise bias memory for the final expression, as suggested above in the discussion of RM boundary conditions and averaging.

What is unclear is why averaging should win out over anticipation in these particular studies. An important next step would seem to be to design experiments, preferably using naturalistic sequences, in which we can attempt to measure both effects. My hunch is that the need to predict versus the need to maintain a stable estimate of the current state often compete when we interact with any form of dynamic stimuli. The representations that guide our behaviour in a given situation, the product of this competition, are almost certainly highly context sensitive. Our memory for the final expression on a face, for example, may be particularly sensitive to emotional engagement (Palumbo and Jellema, 2013) and response timing (Freyd, 1987), coming to reflect different solutions to the anticipation-averaging conflict depending on the precise experimental scenario.

Acknowledgements

Many thanks to my old colleagues and classmates at the University of Oregon who were good enough to be filmed as part of my PhD work, particularly Diego Fernandez-Duque, whose grad school era smile makes an appearance in this paper. I continue to be very grateful to Jennifer Freyd, my advisor, for her inspiration, supervision and friendship and to my other committee members, Mike Posner, Helen Neville and Russ Tomlin, for their guidance. Finally, thanks to Amy Hayes and Jiri Lukavsky for useful comments on earlier versions of this paper.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn. Sci.* 15(3), 122–131.

Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychol Sci* 12(2), 157–162.

Brehaut, J. C. and Tipper, S. P. (1996). Representational momentum and memory for luminance. *J*

Exp Psychol Hum Percept Perform 22(2), 480–501.

Bruce, V., Doyle, T., Dench, N. and Burton, M. (1991). Remembering facial configurations. *Cognition* 38(2), 109–144.

Bruce, V. and Young, A. (2012). *Face Perception* (1st ed.). Psychology Press.

Courgeon, M., Amorim, M. A., Giroux, C. and Martin, J. C. (2010). Do users anticipate emotion dynamics in facial expressions of a virtual character. In *Proc. 23rd int. conf. comput. animat. soc. agents (casa 2010)*. Saint Malo, France.

de Fockert, J. and Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Q J Exp Psychol* 62(9), 1716–1722.

de Gelder, B., Teunisse, J.-P. and Benson, P. J. (1997). Categorical Perception of Facial Expressions: Categories and their Internal Structure. *Cogn. Emot.* 11(1), 1–23.

Edwards, K. (1998). The Face of Time: Temporal Cues in Facial Expressions of Emotion. *Psychol. Sci.* 9(4), 270–276.

Ekman, P. and Friesen, W. V. (1976). *Pictures of Facial Affect.* Consulting Psychologists Press.

Etcoff, N. L. and Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition* 44(3), 227–240.

Faust, M. E. (1990). *Representational Momentum: A Dual Process Perspective* (Doctoral dissertation, University of Oregon, Eugene, OR, USA).

Finke, R. A. and Shyi, G. C. (1988). Mental extrapolation and representational momentum for complex implied motions. *J Exp Psychol Learn Mem Cogn* 14(1), 112–120.

Freyd, J. J. (1987). Dynamic mental representations. *Psychol. Rev.* 94(4), 427–438.

Freyd, J. J. and Finke, R. A. (1984). Representational momentum. *J. Exp. Psychol. Learn. Mem. Cogn.* 10(1), 126–132.

Freyd, J. J. and Johnson, J. Q. (1987). Probing the time course of representational momentum. *J Exp Psychol Learn Mem Cogn* 13(2), 259–268.

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M. and Prinz, W. (2007). Predicting point-light actions in real-time. *Neuroimage* 36 Suppl 2, T22–32.

Gray, R. and Thornton, I. M. (2001). Exploring the link between time to collision and representational momentum. *Perception* 30(8), 1007–1022.

Haberman, J., Harp, T. and Whitney, D. (2009). Averaging facial expression over time. *J Vis* 9(11), 1.1–13.

Haberman, J. and Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* 17(17), R751–3.

Haberman, J. and Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *J Exp Psychol Hum Percept Perform* 35(3), 718–734.

Hubbard, T. L. (1993a). Auditory representational momentum: Musical schemata and modularity. *Bull. Psychon. Soc.* 31(3), 201–204.

Hubbard, T. L. (1993b). The effect of context on visual representational momentum. *Mem Cogn.* 21(1), 103–114.

Hubbard, T. L. (1996). Displacement in depth: representational momentum and boundary extension. *Psychol Res* 59(1), 33–47.

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychon Bull Rev* 12(5), 822–851.

Hubbard, T. L. and Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Percept Psychophys* 44(3), 211–221.

Jellema, T., Pecchinenda, A., Palumbo, L. and Tan, E. G. (2011). Biases in the perception and affective valence of neutral facial expressions induced by the immediate perceptual history. *Vis. cogn.* 19(5), 616–634.

Johnston, H. M. and Jones, M. R. (2006). Higher order pattern structure influences auditory representational momentum. *J Exp Psychol Hum Percept Perform* 32(1), 2–17.

Kerzel, D. (2003a). Attention maintains mental extrapolation of target position: irrelevant distractors eliminate forward displacement after implied motion. *Cognition* 88(1), 109–131.

Kerzel, D. (2003b). Mental extrapolation of target position is strongest with weak motion signals and motor responses. *Vis. Res.* 43(25), 2623–2635.

Knappmeyer, B., Thornton, I. M. and Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vis. Res.* 43(18), 1921–1936.

Krumhuber, E. and Kappas, A. (2005). Moving Smiles: The Role of Dynamic Components for the Perception of the Genuineness of Smiles. *J. Nonverbal Behav.* 29(1), 3–24.

Marian, D. E. and Shimamura, A. P. (2013). Contextual influences on dynamic facial expressions. *Am J Psychol* 126(1), 53–65.

O'Toole, A. J., Roark, D. A. and Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *J. Vis.* 2(7), 604–604.

O'Toole, A. J. and Roark, D. (2010). Memory for Moving Faces: The Interplay of Two Recognition Systems. *Dyn. Faces Insights from Exp. Comput.*

Palumbo, L. and Jellema, T. (2013). Beyond face value: does involuntary emotional anticipation shape the perception of dynamic facial expressions? *PLoS One* 8(2), e56003.

Pilz, K. S., Thornton, I. M. and Bülthoff, H. H. (2005). A search advantage for faces learned in motion. *Exp Brain Res* 171(4), 436–447.

Posner, M. I. and Keele, S. W. (1968). On the genesis of abstract ideas. *J Exp Psychol* 77(3), 353–363.

Russell, J. A. and Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *J. Exp. Psychol. Gen.* 116, 223–237.

Sato, W. and Yoshikawa, S. (2004). Brief Report The dynamic aspects of emotional facial expressions. *Cogn. Emot.* 18(5), 701–710.

Schmidt, K. L., Cohn, J. F. and Tian, Y. (2003). Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles. *Biol Psychol* 65(1), 49–66.

Srismith, D., Oxner, M., Hayward W, G. and Thornton I, M. (2014). An other-race bias when processing groups of faces. *Perception* 43.

Thornton, I. M. (1997). *The perception of dynamic human faces* (Doctoral dissertation, University of Oregon, Eugene). Available to download at http://www.ianthornton.com/publications/.

Thornton, I. M. and Hayes, A. (2004). Anticipating action in complex scenes. *Vis. cogn.* 11(2-3), 341–370.

Thornton, I. M. and Kourtzi, Z. (2002). A matching advantage for dynamic human faces. *Perception* 31(1), 113–132.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res* 168(3), 242–249.

Uono, S., Sato, W. and Toichi, M. (2009). Brief Report: Representational Momentum for Dynamic Facial Expressions in Pervasive Developmental Disorder. *J Autism Dev Disord* 40(3), 371–377.

Uono, S., Sato, W. and Toichi, M. (2014). Reduced representational momentum for subtle dynamic facial expressions in individuals with autism spectrum disorders. *Res. Autism Spectr. Disord.* 8(9), 1090–1099.

Verfaillie, K. and Daems, A. (2002). Representing and anticipating human actions in vision. *Vis. cogn.* 9(1-2), 217–232.

Verfaillie, K. and D'Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *J Exp Psychol Learn Mem Cogn* 17(2), 302–313.

Verfaillie, K., Troy, A. D. and Rensbergen, J. V. (1994). Transsaccadic integration of biological motion. *J Exp Psychol Learn Mem Cogn* 20(3), 649–670.

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P. and Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Front Psychol* 5, 633.

Yoshikawa, S. and Sato, W. (2008). Dynamic facial expressions of emotion induce representational momentum. *Cogn Affect Behav Neurosci* 8(1), 25–31.